Toward a Unified Framework for Point Set Registration

Feiran Li, Kent Fujiwara and Yasuyuki Matsushita

***Abstract*— Point set registration plays a critical role in robotics and computer vision. Early methods considered registration as a purely geometric problem, presenting excellent extensibility for various tasks due to their explicit handling of correspondences; statistical methods were later introduced to handle noise. However, the two categories of algorithms have evolved independently without sharing much in common. In this paper, we leverage the concept of information geometry to theoretically unify the two classes together by interpreting them as the same operation but in different spaces associated with respective metrics. Moreover, based on the proposed unification, we also develop a novel bandwidth estimation strategy to solve the long-standing problem of statistical registration algorithms, and demonstrate its theoretical and practical advantages over deterministic annealing, the most commonly used technique. We also present a case study to show how geometric and statistical approaches can benefit from each other.**

I. Introduction

Pairwise point set registration is a central problem in computer vision and graphics, which plays an important role in 3D reconstruction, shape retrieval, medical imaging, and robot vision. It is formulated as finding the optimal rigid transformation $\mathbf{T}$, consisting of rotation $\mathbf{R}$ and translation $\mathbf{t}$, between two point sets $A = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ and $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$.

The difficulty of registration lies in the unknown correspondences between $\mathbf{b}_j$ and $\mathbf{a}_i$. As pioneering approaches, geometric algorithms such as Iterative Closest Point (ICP) [1] first proposed to tackle this problem by alternating between establishing correspondences and optimizing $\mathbf{T}$. Other methods based on statistics [2], [3], graph matching [4], feature correspondence [5], and deep learning [6] were later introduced to address various challenges. In this work, we focus on the geometric and statistical approaches since they are used by a vast majority of registration tasks.

While statistical methods excel at coping with various types of noise, geometric ones provide more choices in various conditions due to their explicit handling of correspondences. Despite the maturity of these methods, there still exist some long-standing problems. First, the two kinds of algorithms evolved independently. As a result, the relationship between the two classes is unclear, preventing them from sharing techniques for mutual benefit. Second, existing statistical registration methods lack effective strategies to estimate the bandwidths, and heavily depend on empirical tuning without any theoretical justifications [7]. Consequently, their performance is hindered from further enhancement.

In this work, we aim at developing a unified framework as well as a reasonable bandwidth estimation strategy to tackle the aforementioned problems. For unification, we show that the relation between geometric and statistical registration algorithms can be explained from an information geometric perspective, where both geometric and statistical approaches are regarded as conducting the same operation but in different spaces. For bandwidth estimation, we show that the objective function developed for registration intrinsically serves as a guidance for estimating bandwidths; hence we do not need to tune them empirically. We summarize our contributions as follows:

- A theoretical unification that bridges the geometric and statistical registration approaches for mutual benefit.
- A bandwidth estimation strategy with theoretical and practical advantages over existing approaches.

It is noteworthy that there already exist some attempts to bridge the geometric approaches with the statistical ones via expectation-maximization (EM) [7], [8]. However, the fact that the underlying metric of the EM algorithm is the Kullback-Leibler divergence (KL-divergence) prevents these works from explaining other divergence-minimization-based methods [9], [10]. As a result, such a unification cannot generalize to all the cases. On the other hand, our explanation suits all cases well without exception.

II. Related work

Geometric and statistical methods possess their respective merits in solving the point set registration problem. Specifically, owing to the explicit handling of correspondences, geometric methods are more flexible in tasks such as symmetric registration and outlier rejection. On the other hand, statistical ones are more accurate under conditions such as point-wise varying noises or clustering-based registration. We summarize their common objectives in Table I, where MM and KDE respectively stand for mixture model (*a.k.a.*, mixture family distribution) and kernel density estimation.

a) Geometric approaches for registration: By formulating the objective as the orthogonal Procrustes problem [11] with unknown correspondences, ICP [1] leads the development of geometric approaches for point set registration. Some variants of it present alternatives on establishing correspondences. For example, EM-ICP [7] relaxes the one-to-one correspondence rule to soft assignment to deal with inaccurate initialization. Gold *et al.* [12] convert the soft assignment matrix into a doubly stochastic one to allow symmetric registration. Trimmed-ICP [13] robustifies the

*This research was supported by NII CRIS collaborative research program operated by NII CRIS and LINE Corporation.

* Feiran Li and Yasuyuki Matsushita are with Osaka University. Kent Fujiwara is with LINE Corporation.

* {li.feiran, yasumat}@ist.osaka-u.ac.jp, kent.fujiwara@linecorp.com

TABLE I: Summary of geometric and statistical registration algorithms.

	Problem formulation	Example
Geometric	Orthogonal Procrustes problem w/ unknown correspondence	ICP [1]
Statistical	Distribution-fitting of MM	CPD [2]
Statistical	Divergence-minimization between MM/KDE	GMMReg [3]

standard ICP algorithm by trimming out probable mis-correspondences. There also exist other extensions that study global optimization and efficiency. For instance, Yang *et al*. [14] propose a Branch-and-Bound search strategy to guarantee global optimization. Greenspan and Yurick [15] use the approximated k-d tree for reducing the computational cost of matching. The point-to-plane [16], [17] and plane-to-plane [18] variants substitute the point-to-point correspondence for faster convergence and symmetric registration.

b) Statistical approaches for registration: Statistical methods treat point sets as mixture family distributions to cope with noise. Some works consider it as a distribution-fitting problem. For example, Myronenko *et al*. [2] consider one point set as a Gaussian mixture model (GMM) and another one as samples drawn from it. Consequently, it casts the registration problem to estimating the parameters of the distribution, which is solvable via the EM algorithm. Another thread of work fits MMs to both point sets, and the optimal transformation can be estimated by minimizing the divergence between them. For example, Jian and Vemuri [8] recommend minimizing the *L2*-divergence between GMMs. Other distributions and divergences are also explored, such as Student's *t*-distribution [19], hybrid mixture model [20], Jensen-Shannon divergence [10] and Cauchy-Schwarz divergence [9].

c) Attempts to unify geometric and statistical approaches: There already exists literature that attempts to unify geometric and statistical approaches. For example, Granger and Pennec [7] interpret ICP as minimizing the sum of Mahalanobis distances between Gaussian distributions. Jian and Vemuri [8] explain ICP as an approximation of minimizing KL-divergence between two GMMs. Segal *et al*. [18] use maximum likelihood estimation (MLE) to maximize the sum of probabilities of distances between paired points, which actually leads to the same conclusion drawn by [8] but from a different perspective. The main difference between these existing works and our proposal is that, while they treat ICP as a particular case of their respective objective functions without referring to its variants, we provide thorough theoretical explanations to unify geometric and statistical methods without any exception.

III. A UNIFIED FRAMEWORK FOR GEOMETRIC AND STATISTICAL APPROACHES

We first need to unify distribution-fitting and divergence-minimization methods within the statistical class before unifying the geometric and the statistical ones.

Remark 1: The procedure of maximizing likelihood is equivalent to minimizing KL-divergence: $\underset{\theta}{\operatorname{argmax}}\, P\left(x;\theta\right) = \underset{\theta}{\operatorname{argmin}}\, \mathtt{KL}\left(P\left(x;\theta^{\star}\right)|P\left(x;\theta\right)\right)$; where θ is the estimated parameter and $\theta^{\star}$ is the ground-truth one.

Per Remark 1, the objective function of distribution-fitting methods can be rewritten as minimizing divergence:

$$\begin{aligned}&\underset{\mathbf{T}}{\operatorname{argmax}} \prod P\left(\mathbf{b}_j; A, \mathbf{T}, \theta\right) \equiv \\ &\underset{\mathbf{T}}{\operatorname{argmin}}\, \mathtt{KL}\left(P\left(B;\theta_b\right)|P\left(A;\mathbf{T},\theta_a\right)\right),\end{aligned} \tag{1}$$

where $P\left(A;\mathbf{T},\theta_a\right) = \sum \alpha_i P\left(\mathbf{a}_i;\mathbf{T},\theta_{ai}\right)$ and $P\left(B;\theta_b\right) = \sum \beta_j P\left(\mathbf{b}_j;\theta_{bj}\right)$ are MMs constructed from point sets A and B. We hereafter denote them as $P\left(A\right) = \sum \alpha_i P\left(\mathbf{a}_i\right)$ and $P\left(B\right) = \sum \beta_j P\left(\mathbf{b}_j\right)$ for notational simplicity. Now, all the statistical methods can be seen as a divergence-minimization problem, which lays the foundation for our unification.

KL-divergence is not the only choice for formulating the objective function. In fact, as employed by numerous related works [8]–[10], there are many off-the-shelf divergences for selection. However, not all of them are suitable for the registration problem. As a naive rule, we desire the selected divergence to be invariant when the points are encoded in different ways. Specifically, given a mapping h, we want the divergence to satisfy

$$D\left(P|Q\right) = D\left(h\left(P\right)|h\left(Q\right)\right). \tag{2}$$

Eq. (2) directly leads to the theory of f-divergence, which contains a group of famous divergences such as KL-divergence, total variation, and all of the α-divergences. We present its formal definition in Lemma 1.

Lemma 1: The f-divergence: $D_f\left(P|Q\right) \triangleq \int_{\Omega} pf\left(\frac{q}{p}\right)d\mu$ is an invariant measurement, where μ is the reference distribution satisfying $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$, and f is a smooth convex function. It is convex w.r.t. both P and Q.

A detailed introduction together with proof of Lemma 1 can be found in *P54-P65* of [21]. Although the f-divergence is a reasonable choice to formulate the objective function, it is seldom analytically computable when the incorporated distributions are MMs. Therefore, we need to find a solvable upper bound for optimization purposes.

Given two MMs written in the form of $P = \sum_i^M \alpha_i P_i$ and $Q = \sum_j^N \beta_j Q_j$, where P_i and Q_j are the respective component distributions of P and Q, and α_i and β_j are their weights; we can treat them as having the same number of components

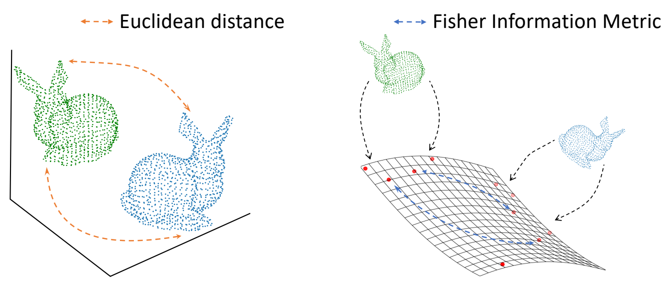
$$\begin{aligned}&\begin{cases} P = \sum_i^{MN} \alpha_{i,n} P_{i,n} \\ Q = \sum_j^{MN} \beta_{j,m} Q_{j,m} \end{cases}, \\ &\text{where } \begin{cases} \alpha_{i,1} = \cdots = \alpha_{i,n} = \frac{\alpha_i}{N} \\ \beta_{j,1} = \cdots = \beta_{j,m} = \frac{\beta_j}{M} \\ P_{i,1} = \cdots = P_{i,n} = P_i \\ Q_{j,1} = \cdots = Q_{j,m} = Q_j \end{cases}.\end{aligned} \tag{3}$$

Algorithm 1: The unified algorithm for point set registration

Input: Point sets A and B, initial transformation $\mathbf{T}$

While not converged:

- **Step 1**: Find a mapping $\varepsilon : \mathbb{R}^3 \rightarrow \mathcal{M}$ (or $\mathbb{R}^3$) for each point $\mathbf{a}_i$ and $\mathbf{b}_j$ under current $\mathbf{T}$.
- **Step 2**: Establish correspondences on $\mathcal{M}$ (or $\mathbb{R}^3$) under current $\mathbf{T}$.
- **Step 3**: Minimize the sum of divergences (or Euclidean distances) w.r.t. $\mathbf{T}$.



(a) Geometric (b) Statistical

Fig. 1: Conceptual illustration of how geometric and statistical methods work. While geometric approaches operate on Euclidean space, statistical ones work on statistical manifold.

Owing to its convexity, we can define an upper bound of the f-divergence $D_f\left(P|Q\right)$ in the summation form via Jensen's inequality [22]

$$\begin{aligned} D_f\left(P|Q\right) &= D_f\left(\sum_{i,n}^{M,N}\alpha_{i,n}P_{i,n}|\sum_{j,m}^{M,N}\beta_{j,m}Q_{j,m}\right) \\ &\leqslant \sum_{i,n,j,m}^{M,N,M,N} D_f\left(\alpha_{i,n}P_{i,n}|\beta_{j,m}Q_{j,m}\right). \end{aligned} \tag{4}$$

Moreover, since MMs are invariant to permutation of indices, we can always tighten the upper bounds in the form of

$$\begin{aligned} D_f\left(P|Q\right) &\leqslant \sum_{i,n,j,m}^{M,N,M,N} D_f\left(\alpha_{i,n}P_{i,n}|\beta_{s(i)}Q_{s(i)}\right) \\ &\leqslant \sum_{i,n,j,m}^{M,N,M,N} \pi_{ij}D_f\left(\alpha_{i,n}P_{i,n}|\beta_{j,m}Q_{j,m}\right) \\ &= \sum_{i,n,j,m}^{M,N,M,N} \pi_{ij}D_f\left(\frac{\alpha_i}{N}P_{i,n}|\frac{\beta_j}{M}Q_{j,m}\right), \end{aligned} \tag{5}$$

where $\beta_{s(i)}Q_{s(i)}$ denotes the nearest component of Q w.r.t. $\alpha_{i,n}P_{j,m}$, and π_{ij} is a weight coefficient *s.t.* $\sum_j \pi_{ij} = 1$.

Since the MMs generated by the two point sets commonly obey $\alpha_i = \beta_j = \frac{1}{M}$, by applying Eq. (5) to the registration problem, the objective function can be written as

$$J = \sum_{i,n,j,m}^{M,N,M,N} \pi_{ij}D_f\left(\frac{1}{MN}P\left(\mathbf{a}_i\right)|\frac{1}{NM}P\left(\mathbf{b}_j\right)\right). \tag{6}$$

Moreover, according to the invariance property of the f-divergence, we can equally scale each component distribution by MN without changing the objective function. Consequently, the above equation is simplified to

$$J = \sum_{i,j}^{M,N} \pi_{ij}D_f\left(P\left(\mathbf{a}_i\right)|P\left(\mathbf{b}_j\right)\right). \tag{7}$$

Recalling that the objective functions of ICP in its generalized form (*i.e.*, soft assignment) is written as

$$J = \sum \pi_{ij}\left\|\mathbf{T}\circ\mathbf{a}_i - \mathbf{b}_j\right\|_2^2, \tag{8}$$

we can see that it is very similar to Eq. (7), except that the Euclidean distance is replaced by the f-divergence. To have a more comprehensive understanding of what statistical methods actually do, we introduce another property of the f-divergence shown in Lemma 2.

Lemma 2: The f-divergence naturally introduces geometric structures to P and Q, where fisher information metric (FIM) is the associated Riemannian metric. Also, we can always use a convex function f satisfying $f(1) = 0$, $f'(1) = 0$ and $f''(1) = 1$ to formulate a standard f-divergence.

Lemma 2 in fact leads to the theory of information geometry [21]. It studies statistics with geometric tools by treating distributions as points on a Riemannian manifold $\mathcal{M}$. For proof:

Proof: Given two neighboring points $P\left(x;\theta\right)$ and $P\left(x;\theta+\delta\theta\right)$ (hereafter denoted as $P\left(\theta\right)$ and $P\left(\theta+\delta\theta\right)$ for notational simplicity) on $\mathcal{M}$, we can obtain the local form of the f-divergence:

$$D_f\left(P\left(\theta\right)|P\left(\theta+\delta\theta\right)\right) = \int P\left(\theta\right)f\left(\frac{P\left(\theta+\delta\theta\right)}{P\left(\theta\right)}\right). \tag{9}$$

By expanding the term $f(\cdot)$ with Taylor expansion to the 2^{nd} order and using the properties $f(1) = 0$, $f'(1) = 0$ and $f''(1) = 1$, Eq. (9) can be rewritten into

$$\begin{aligned} D_f &= \int P\left(\theta\right)\left(f\left(\frac{P\left(\theta\right)}{P\left(\theta\right)}\right) + \frac{\partial f}{\partial\theta}\delta\theta + \frac{1}{2}\delta\theta^T\frac{\partial^2 f}{\partial\theta^2}\delta\theta\right) \\ &= \frac{1}{2}\int P\left(\theta\right)\cdot\left(0 + 0 + \delta\theta^T\frac{\partial^2 f}{\partial\theta^2}\delta\theta\right) \\ &= \frac{1}{2}\delta\theta^T\left(\int\cdot\left(\frac{1}{P\left(\theta\right)}\right)\cdot\left(\frac{\partial}{\partial\theta}P\left(\theta\right)\right)\left(\frac{\partial}{\partial\theta}P\left(\theta\right)\right)\right)\delta\theta \\ &= \frac{1}{2}\delta\theta^T\left(\int P\left(\theta\right)\frac{\partial\log P\left(\theta\right)}{\partial\theta}\frac{\partial\log P\left(\theta\right)}{\partial\theta}\right)\delta\theta \\ &= \frac{1}{2}g_{ij}d\theta^i d\theta^j, \end{aligned} \tag{10}$$

where Einstein Summation form is used and $g_{ij} = \int P\left(\theta\right)\frac{\partial\log P(\theta)}{\partial\theta_i}\frac{\partial\log P(\theta)}{\partial\theta_j}$ is exactly the FIM. ■

Since the last line of Eq. (10) is a general notation for distance, we can also express the euclidean distance in the same form by changing g_{ij} to the Kronecker-delta. Therefore, we can now express the objective functions of both statistical and geometric methods in the unified form of

$$J = \frac{\pi_k}{2} g_{ij}^k d\theta_k^i d\theta_k^j, \tag{11}$$

where k denotes the indices of pairs of points; and g_{ij} encodes distinct metrics for statistical and geometric methods.

With this unified objective function, we can conclude that the two classes of algorithms, in fact, conduct the same operations in different spaces associated with different metrics. Specifically, as illustrated in Fig. 1, they all actually do ICP (or its variants) but in distinct spaces.

One merit of the unified explanation is that, we are freed from having to select between geometric and statistical registration algorithms. *I.e.*, we can always develop an algorithm that imitates ICP and its variants on handling correspondences while keeping the algorithm statistical to handle noise. We summarize the unified algorithm in Algorithm 1, where ε is an endomorphism ($\mathbb{R}^3 \to \mathbb{R}^3$) for geometric methods and a homomorphism ($\mathbb{R}^3 \to \mathcal{M}$) for statistical counterparts.

IV. An intrinsically contained method for bandwidth estimation strategy

We now put Algorithm 1 into practice. A major problem faced by many divergence-minimization-based approaches lies in how to define an effective mapping $\varepsilon : \mathbb{R}^3 \to \mathcal{M}$. Fortunately, the aforementioned f-divergence intrinsically contains an approach for doing so. To maintain consistency with other registration literature, and without loss of generality, we hereafter assume that all the points within a given point set are contaminated by the same isotropic Gaussian noise.

Since the expectation of each Gaussian distribution can be naturally determined as the point position in the Euclidean space, the only remaining problem related to the mapping pertains to the covariances $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$. In the case of clustering-based approaches [23], [24], they are easy to estimate and can be treated as known parameters. However, in general cases, the mapping is considered as a bijection, which leads to the bandwidth estimation problem of KDE.

Empirically, the bandwidth of a general KDE can be estimated beforehand and fixed with rule-of-thumb methods [25]. However, in the context of the registration problem, a dynamically updated one would be more suitable since the relative position of the two point sets varies according to the transformation $\mathbf{T}$.

The heuristic method of deterministic annealing is a popular tool for setting the bandwidths $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$ [3], [8], [9]. Its philosophy is based on the asymptotic behaviors as studied by Chui *et al.* [26]. In detail, $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$ control the correspondences π_{ij} between point sets. The correspondences are vague when $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$ are large, and become asymptotically one-to-one when they approach $\mathbf{0}$. Therefore, in order to take advantage of both soft assignment and one-to-one correspondences to avoid local minima and obtain accurate registration, deterministic annealing gradually decreases $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$ from large initial values until convergence. Although promising performances have been observed, we hesitate to do so since bandwidth estimation is an independent problem in statistics with more effective solutions.

We claim that the objective function developed for registration intrinsically acts as a guidance for dynamically setting the bandwidths. Specifically, it reminds us of the bandwidth estimation technique based on risk functions [27].

Remark 2: A risk function L is defined as the expectation of a loss f: $L = E\left[f\right]$. Applying to bandwidth estimation of KDE, the optimal bandwidth of a KDE p can be retrieved by minimizing $E\left[f\left(p, q\right)\right]$, where q is the underlying ground-truth distribution for reference.

According to the definition in Remark 2, the f-divergence in fact naturally formulates one risk function in the form of $D_f = E_p\left[f\left(\frac{q}{p}\right)\right]$. Therefore, if we alternately fix the point $P\left(\mathbf{a}_i\right)$ or $P\left(\mathbf{b}_j\right)$ as the reference, the f-divergence also serves as an objective function for estimating the bandwidth of the other one.

If the reference distribution q is the ground-truth, a plausible bandwidth of the variable distribution p would be the one that minimizes the difference between them. However, for the registration problem, since q is just temporarily fixed without any guarantees to be faithful, directly tuning p towards it would unavoidably introduce large variance for smoothing purposes. This problem is known as the bias-variance-tradeoff, and we employ the minimizing entropy criterion proposed by Jiang *et al.* [28] as a solution. Specifically, we can limit the variance of a KDE by adding in a regularization term of its entropy, which leads to our final objective function:

$$\begin{aligned} J\left(\mathbf{T}, \mathbf{\Sigma}_a, \mathbf{\Sigma}_b\right) = \sum \pi_{ij} D_f\left(P\left(\mathbf{a}_i\right) | P\left(\mathbf{b}_j\right)\right) \\ + \lambda_1 H\left(P\left(A\right)\right) + \lambda_2 H\left(P\left(B\right)\right), \end{aligned} \tag{12}$$

where $H\left(\cdot\right)$ is the entropy, and λ_1 and λ_2 are the regularization parameters. As commonly used in other registration algorithms, Eq. (12) can be facilely minimized via alternating optimization, which update $\mathbf{T}$, $\mathbf{\Sigma}$ and π alternately.

V. How unification helps mutual benefit: A case study

Based on the aforementioned unification, we can always design an algorithm that takes advantage of both statistical and geometric methods, namely, an algorithm based on divergence-minimization with explicitly maintained correspondences. In this section, we take the symmetric registration task as an example to demonstrate its effectiveness.

A. Problem formulation

Assuming that the two point sets consist of a similar amount of points and no preliminary information of the registration order ($A \to B$ or $B \to A$) is given, symmetric registration aims at alleviating the effects of registration order on the final results. That is, we want $A \to B$ and $B \to A$ to return nearly the same estimations.

This problem is easy to solve for geometric methods. For example, as used in [12], [29], [30], it is a common practice to convert the aforementioned soft assignment matrix π to an (asymptotic) doubly stochastic one in each iteration. In detail, the property of doubly stochastic matrices that each row and column sums to 1 helps to ignore the normalization

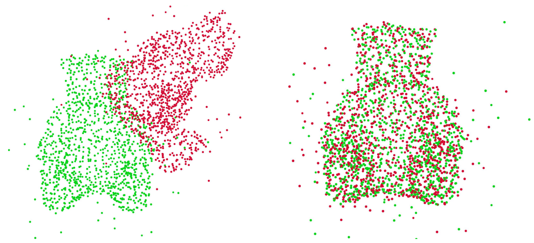


Fig. 2: An example of the synthetic data contaminated by Gaussian noise and outliers. Left: point clouds in their initial poses. Right: Our registration result.

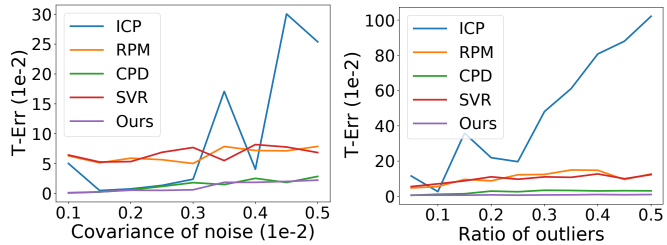


Fig. 3: Comparison w.r.t. varying levels of noise and ratios of outliers. Each experiment consist of 10 different setups.

TABLE II: Errors on synthetic data (both rotational and transformational errors are in $1e-2$ scale).

	Mean			Median		
	$\mathbf{R}$ err	$\mathbf{t}$ err	$\mathbf{T}$ err	$\mathbf{R}$ err	$\mathbf{t}$ err	$\mathbf{T}$ err
ICP	39.49	6.93e-4	39.49	2.91	6.71e-4	2.91
RPM	4.92	7.04e-4	4.93	4.29	6.96e-4	4.29
CPD	1.67	**3.30e-4**	1.67	1.32	**3.24e-4**	1.32
SVR	11.73	6.02e-4	11.73	4.21	4.08e-4	4.21
Ours	**0.93**	3.73e-4	**0.93**	**0.96**	3.56e-4	**0.96**

axis of π, hence the registration order is perfectly balanced. On the other hand, dealing with this task is a weakness of existing statistical methods. In detail, the "A as model and B as data" principle of distribution-fitting-based methods [2], [20] naturally requires a pre-defined registration order. For the divergence-minimization-based methods, although we can rely on some symmetric divergences [9], [10], there still lacks a unified solution when using general divergences.

Our unification helps the statistical algorithms to strengthen symmetric registration. In detail, if we treat each component distribution as a point on $\mathcal{M}$, we can naturally follow the geometric methods and convert the assignment matrix π to doubly stochastic in each iteration. The only difference is that, while geometric methods use the Euclidean distance, the statistical ones must replace it with a divergence.

B. Experiments

We select 4 other candidates for comparison: ICP [1], RPM [12], CPD [2] and SVR [31]. Among them, ICP solves the registration problem in a pure geometric manner. RPM converts the correspondences of ICP to doubly stochastic to realize symmetric registration. CPD and SVR respectively represent the distribution-fitting and divergence-minimization-based statistical approaches. Regarding the implementations used in the experiments, ICP and RPM are implemented by ourselves, CPD is provided by the authors, and SVR is from the *ProbReg* Python library[1]. For all of the following experiments, we assume the datasets are not under extreme conditions (*i.e.*, severe outliers, low overlaps, or significant distinction between numbers of points) since they are commonly considered as independent tasks addressed by extending the 4 base approaches with other techniques [13], [32]–[34].

[1]ProbReg: `http://probreg.readthedocs.io/`. Last accessed on March 24, 2021.

We use the symmetrized KL-divergence to specify the f-divergence for maintaining symmetry. Also, in order to show that the tools developed for geometric methods can be directly applied to statistical ones, we follow RPM [7] and use softmax for soft assignment and Sinkhorn iteration [35] for converting the assignment matrix π to doubly stochastic. For each pair of point sets in both synthetic and realistic datasets, we conduct the registration from both $A \to B$ and $B \to A$. We use $\mathbf{R}-\text{Err} = \left\|\mathbf{I} - \mathbf{R}_{\text{gt}}^T \mathbf{R}_{\text{esti}}\right\|_F$, $\mathbf{t}-\text{Err} = \|\mathbf{t}_{\text{esti}} - \mathbf{t}_{\text{gt}}\|_2$ and $\mathbf{T}-\text{Err} = \left\|\mathbf{I} - \mathbf{T}_{\text{esti}}^{-1} \mathbf{T}_{\text{gt}}\right\|_F$ as the respective error metrics for rotation, translation and transformation. For parameter setup, the regularization coefficients for bias-variance-tradeoff in our method are set to $\lambda_1 = \lambda_2 = 50$. The annealing parameter is set to 0.95 for RPM and 0.9 for SVR, with which we observe better performances. For CPD, the weight of uniform distribution is set to 0.1. The maximum iteration is set to 100 for all the algorithms except SVR, which is set to 10 for its slow clustering procedure. In fact, we observe that once the SVR algorithm succeeds, the number of iterations shows minor effects on its final accuracy.

a) Tests on synthetic data: We first test how different algorithms perform under noisy conditions. For experimental setup, we base the tests on the bunny and the dragon point sets given in [36], [37], and a teddy from Free3D[2]. They are all down-sampled to around 1000 points for a reasonable runtime, as shown in Fig. 2. We add isotropic Gaussian noises with standard deviation $\boldsymbol{\sigma} = 0.0015\mathbf{I}$ and $\boldsymbol{\sigma} = 0.0025\mathbf{I}$, as well as 10% of outliers to the two point sets respectively. The registration is repeated 20 times on each pair of point sets. For each trial, a random ground-truth rotation $\mathbf{R}$ drawn from Euler angles in the range of $(0°, 60°)$ along each axis, and a random translation is used. The results are shown in Table II. As we can see, our statistical symmetric registration method can champion both the rotational and transformational estimations, although it performs negligibly inferior to CPD on estimating the translations.

We also test the robustness of our algorithm w.r.t. different ratios of outliers and levels of noises. For setup, we use the aforementioned bunny point set and fix the ground-truth rotation to $30°$ in Euler angles. For outliers, we use the 10 ratios uniformly distributed within 5% to 50%. For noise, we vary the covariances of Gaussian noise from $1e-3\mathbf{I}$ to $5e-3\mathbf{I}$ with a step size of $5e-4\mathbf{I}$. We repeat the experiment

[2]Free3D `http://www.free3d.com`. Retrieved March 24, 2021.

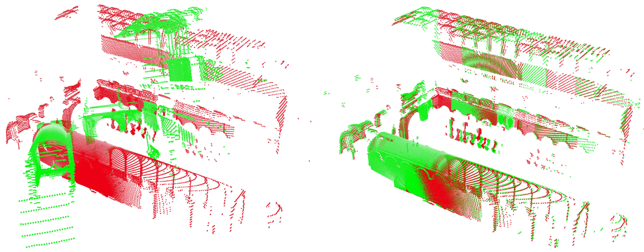


Fig. 4: An example of the ETH Hauptgebaude dataset. Some points are cropped for better visualization. Left: the point sets in their initial poses. Right: our registration result.

TABLE III: Errors on the laser scan data.

	Mean			Median		
	$\mathbf{R}$ err	$\mathbf{t}$ err	$\mathbf{T}$ err	$\mathbf{R}$ err	$\mathbf{t}$ err	$\mathbf{T}$ err
ICP	2.11e-2	0.61	0.61	1.90e-2	0.55	0.61
RPM	5.06e-2	**0.11**	**0.11**	4.65e-2	**0.11**	**0.11**
CPD	2.08e-2	0.15	0.15	1.93e-2	0.15	0.15
SVR	6.69e-2	0.48	0.49	4.21e-2	0.36	0.49
Ours	**1.66e-2**	0.12	0.12	**1.56e-2**	0.12	0.12

10 times with each ratio and each level. The mean errors are plotted in Fig. 3. As shown, for the case of different levels of noise, our proposal in general performs better than CPD. For the case of distinct ratios of outliers, our method can stably present higher accuracy compared with the others.

b) Tests on real data: We also conduct experiments on laser scan dataset to study the real-world performances. In detail, we base the tests on the first 10 scans of the ETH Hauptgebaude dataset [38], as shown in Fig. 4. This dataset is challenging as it consists of repetitive elements. Since some of the original ground-truth rotations are either extremely large that none of the candidates can give reasonable estimations, we manually reset the rotations to the range of $(0°, 60°)$ in Euler angles along each axis. We also change the weight of the uniform distribution of CPD to 0 as we found it barely succeeds with other numbers. Again, each point set is down-sampled to around 1000. The results are shown in Fig. III. As can be seen, our method outperforms all the others on estimating rotations. For translational and transformational errors, although RPM presents slightly better results, the relative differences between them and ours are small compared to the improvements on rotations.

VI. Performance of the Proposed Bandwidth Estimation Strategy

In this section, we study the performance of our proposed bandwidth estimation strategy. For experimental setup, we use the aforementioned synthetic bunny with a fixed ground-truth rotation of $(30°, 30°, 30°)$ in Euler angles; and keep noise, outliers, and parameter settings the same as mentioned above.

a) Comparisons between deterministic annealing and our proposal: We study the relative performance between the commonly used deterministic annealing technique and our proposal to show the superiority of our bandwidth estimation strategy. Specifically, given initial bandwidths,

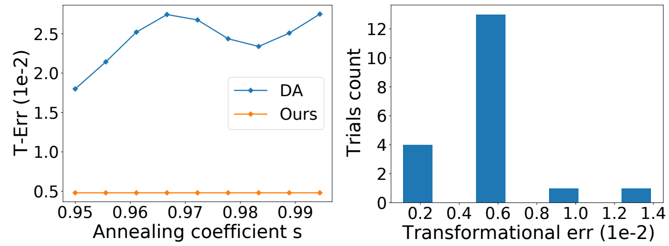


Fig. 5: Comparison of deterministic annealing and our bandwidth estimation strategy. Left: Accuracy w.r.t. different annealing parameters. The orange line is achieved by fixing the tradeoff parameters to $\lambda_1 = \lambda_2 = 50$. Right: Error distribution with $\lambda_1 = \lambda_2$ uniformly drawn from $[1, 60]$.

the deterministic annealing technique multiplies it with a shrinkage coefficient s after each iteration. In our implementation, we observe that the registration performances become significantly unstable when s is within the range of $(0.9, 0.95)$, and it hardly succeeds when $s \in (0, 0.9)$. Therefore, we limit s to be drawn from $(0.95, 1)$ to make the comparison meaningful. The results are reported in Fig. 5. As shown in the figure, the performance of deterministic annealing is not stable as minor changes on s led to conspicuous shifts in the estimated results. Moreover, its accuracy is still significantly inferior to our strategy.

b) Stability w.r.t. different λ: As the only free parameters in our proposal, it is meaningful to study how different λ_1 and λ_2 for bias-variance-tradeoff can affect the performance. We again simply assume λ_1 and λ_2 to share the same value: $\lambda_1 = \lambda_2 = \lambda$. For testing purposes, we uniformly pick 20 trials with $\lambda \in [1, 60]$. The final errors are summarized in the histogram in Fig. 5. As can be seen, our bandwidth estimation strategy is stable w.r.t. varying λ within a broad tuning range. In detail, the apparent dominant error suggests that a large scope of different λ can eventually lead the algorithm to the same estimation. Moreover, compared to the performance of the deterministic annealing mentioned above, the difference between the best and worst estimations of our proposal is only in 10^{-3} magnitude.

VII. Discussion and Conclusion

In this work, we present a unified framework by explaining geometric and statistical point set registration algorithms as conducting the same operations in different spaces. We also derive a bandwidth estimation strategy for general divergence-minimization-based methods, which demonstrates its effectiveness over the existing heuristic method. Also, by taking the symmetric registration task as an example, we show how to combine the strengths of both classes, whose effectiveness is demonstrated in various experiments.

For future work, we plan to combine the statistical registration methods with other geometric ones to deal with cases such as partial overlaps and sparse-to-dense registration. Another direction lies in exploring how to select a suitable f-divergence for the registration problem, which remains an open problem in information science.

References

[1] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 14, pp. 239–256, 1992.

[2] A. Myronenko and X. Song, "Point set registration: Coherent point drift," Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 32, no. 12, pp. 2262–2275, 2010.

[3] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 33, no. 8, pp. 1633–1645, 2010.

[4] F. Zhou and F. De la Torre, "Factorized graph matching," in Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 127–134, 2012.

[5] S. Gumhold, X. Wang, and R. S. MacLeod, "Feature extraction from point clouds," in Proceedings of International Meshing Roundtable, Sandia National Laboratories, 2001.

[6] G. Elbaz, T. Avraham, and A. Fischer, "3d point cloud registration for localization using a deep neural network auto-encoder," in Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4631–4640, 2017.

[7] S. Granger and X. Pennec, "Multi-scale em-icp: A fast and robust approach for surface registration," in Proceedings of European Conference on Computer Vision (ECCV), pp. 418–432, 2002.

[8] B. Jian and B. C. Vemuri, "A robust algorithm for point set registration using mixture of gaussians," in Proceedings of International Conference on Computer Vision (ICCV), vol. 2, pp. 1246–1251, 2005.

[9] E. Hasanbelliu, L. S. Giraldo, and J. C. Principe, "Information theoretic shape matching," Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 36, no. 12, pp. 2436–2451, 2014.

[10] F. Wang, B. C. Vemuri, and A. Rangarajan, "Groupwise point pattern registration using a novel cdf-based jensen-shannon divergence," in Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 1283–1288, 2006.

[11] J. C. Gower and G. B. Dijksterhuis, Procrustes problems (Oxford Statistical Science Series Book), vol. 30. 2004.

[12] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2d and 3d point matching: Pose estimation and correspondence," Pattern Recognition, vol. 31, no. 8, pp. 1019–1031, 1998.

[13] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," in Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), vol. 3, pp. 545–548, 2002.

[14] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-icp: A globally optimal solution to 3d icp point-set registration," Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 38, no. 11, pp. 2241–2254, 2015.

[15] M. Greenspan and M. Yurick, "Approximate kd tree search for efficient icp," in Proceedings of International Conference on 3-D Digital Imaging and Modeling, pp. 442–448, 2003.

[16] K.-L. Low, "Linear least-squares optimization for point-to-plane icp surface registration," Tech. Rep. 10, Chapel Hill, University of North Carolina, 2004.

[17] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," Image and Vision Computing, vol. 10, no. 3, pp. 145–155, 1992.

[18] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp," in Proceedings of Robotics: Science and Systems, vol. 2, p. 435, 2009.

[19] Z. Zhou, J. Zheng, Y. Dai, Z. Zhou, and S. Chen, "Robust non-rigid point set registration using student's-t mixture model," PloS one, vol. 9, no. 3, p. e91381, 2014.

[20] Z. Min, J. Wang, and M. Q.-H. Meng, "Robust generalized point cloud registration using hybrid mixture model," in Proceedings of International Conference on Robotics and Automation (ICRA), pp. 4812–4818, 2018.

[21] S. Amari, Information geometry and its applications (Applied Mathematical Sciences Book), vol. 194. 2016.

[22] Y. Qiao and N. Minematsu, "A study on invariance of f-divergence and its application to speech recognition," Transactions on Signal Processing (TSP), vol. 58, no. 7, pp. 3884–3890, 2010.

[23] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in Proceedings of International Conference on Intelligent Robots and Systems (IROS), vol. 3, pp. 2743–2748, 2003.

[24] T. Stoyanov, M. Magnusson, and A. J. Lilienthal, "Point set registration through minimization of the l2 distance between 3d-ndt models," in Proceedings of International Conference on Robotics and Automation (ICRA), pp. 5196–5201, 2012.

[25] B. W. Silverman, Density estimation for statistics and data analysis. 2018.

[26] H. Chui and A. Rangarajan, "A feature registration framework using mixture models," in Proceedings of Workshop on Mathematical Methods in Biomedical Image Analysis, pp. 190–197, 2000.

[27] W. Qiao et al., "Asymptotics and optimal bandwidth for nonparametric estimation of density level sets," Electronic Journal of Statistics, vol. 14, no. 1, pp. 302–344, 2020.

[28] J. Jiang, Y.-L. He, D.-X. Dai, and J. Z. Huang, "A new kernel density estimator based on the minimum entropy of data set," Information Sciences, vol. 491, pp. 223–231, 2019.

[29] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," Journal on Computer Vision and Image Understanding (CVIU), vol. 89, no. 2-3, pp. 114–141, 2003.

[30] Z. J. Yew and G. H. Lee, "Rpm-net: Robust point matching using learned features," in Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[31] D. Campbell and L. Petersson, "An adaptive data representation for robust point-set registration and merging," in Proceedings of International Conference on Computer Vision (ICCV), pp. 4292–4300, 2015.

[32] L. Peng, G. Li, M. Xiao, and L. Xie, "Robust cpd algorithm for non-rigid point set registration based on structure information," PloS one, vol. 11, no. 2, 2016.

[33] J. Zhu, C. Jin, Z. Jiang, S. Xu, M. Xu, and S. Pang, "Robust point cloud registration based on both hard and soft assignments," Optics & Laser Technology, vol. 110, pp. 202–208, 2019.

[34] J. Ma, J. Zhao, J. Jiang, and H. Zhou, "Non-rigid point set registration with robust transformation estimation under manifold regularization," in Proceedings of AAAI Conference on Artificial Intelligence, 2017.

[35] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," Annals of Mathematical Statistics, vol. 35, no. 2, pp. 876–879, 1964.

[36] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in Proceedings of SIGGRAPH, pp. 303–312, 1996.

[37] S. Choi, Q.-Y. Zhou, S. Miller, and V. Koltun, "A large dataset of object scans," arXiv preprint arXiv:1602.02481, 2016.

[38] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, "Challenging data sets for point cloud registration algorithms," International Journal of Robotics Research (IJRR), vol. 31, no. 14, pp. 1705–1711, 2012.